

## Book review

# Genomics and evolution of microbial eukaryotes

**Edited by Laura Katz and Debashish Bhattacharya**

Oxford University Press, Oxford, 2006

Hardcover. 256 pp.

Price: £ 55.00/US\$ 125.00

ISBN 9780198569749

Genome-scale data – complete genome sequences, deep sequencing of cDNA libraries, expression-array data and sequence tags – continue to revolutionise the biological sciences. For increasingly many organisms of interest, eukaryotes as well as prokaryotes, research programs are no longer designed as a stepwise dissection of phenotype through successively more-fundamental strata of pathways, protein sequences, messages, genetic loci and finally genes. Extensive genomic, expression and proteomic data are instead available up-front, and the challenge has become to identify and validate those molecular regulatory and interaction networks most immediately relevant to the question at hand: production of a metabolite or biopolymer, invasion of a host, or exploitation of an ecological niche.

For the evolutionary biologist, genome data are a treasure-trove of evidence on shared ancestry, offering the promise of phylogenies based no longer on the limited and perhaps biased sampling dictated by suitability as targets for PCR primers, but rather on complete sets of exons – or perhaps on comprehensive analyses of protein folds, gene content or gene order. Indeed, comparative genomics promises much more, including a system-wide understanding of the origins and principles of genetic transmission, genome structure, novel and partitioned enzyme function, conserved and unique molecular-interaction networks, multicellularity, development, selection, and population dynamics.

Genomics came somewhat late to microbial eukaryotic biology. The delay came about in part because the DNA-sequencing technologies first trialled on bacteria proved to be highly scalable and so were quickly applied to yeast, then worm and fly, then human, mouse, and the major animals and plants of agricultural interest. But it did not help that the morphologically simpler eukaryotes present such a breathtaking diversity of specialised, hence difficult, genetics and chromosomal biology. When sequencing is relatively costly, resources naturally flow to systems with organised research communities, reference strains well-characterised at the chromosomal and genomic levels, deep DNA and EST libraries, gene-expression resources and experimentally tractable genetics.

Even when genomics arrives, as it has now for some microbial eukaryotes, it presents a moving target. Data become available in stages – ESTs first, then individual

genomic reads, then contigs, scaffolds, and successive builds of chromosome assembly. Annotation is focused first on exon-intron structure, only later on motifs, pathways, co-regulation, and comparative analyses. cDNAs may not be full-length, expression platforms present a maze of hazards, and false-positive predictions abound. More insidiously, sequence data from domesticated strains and clonal populations are welcomed as representing “the” genome, whereas expression and proteomic data intentionally (to maximise sampling of encoded information) reflect a diverse admixture of life-cycle and physiological states.

This book, edited by two outstanding younger researchers who were early adopters of genome-scale technology in application to ciliates and photosynthetic protists respectively, takes aim at this moving target for microbial eukaryotes *circa* mid-2005: the published genome of *Plasmodium falciparum*, plus partial sequences of 11 other *Plasmodium* strains; the partially annotated *Tetrahymena thermophila* genome, plus data from *Paramecium tetraurelia* and *Oxytricha trifallax*; published genomes of *Trypanosoma brucei*, *T. cruzi* and *Leishmania major*; the published but partially assembled genome of *Entamoeba histolytica*; the compact *Encephalitozoon cuniculi* genome, with an initial comparison with that of *Antonospora locustae* (formerly *Nosema locustae*); nucleomorphs of *Bigelowiella natans* and *Guillardia theta*; the published *Thalassiosira pseudonana* genome, and ESTs of *Phaeodactylum tricornutum*; and the published genome of *Dictyostelium discoideum*. Authors additionally had access to ESTs of *Alexandrium tamarense*, *Karenia brevis* and four other dinoflagellates, and GSS and EST sequences from *Spironucleus barkhanus*.

Although released in April 2004, the *Cyanidioschyzon merolae* genome is surprisingly not treated. The April 2004 genome of *Cryptosporidium parvum* is discussed only briefly, while the October 2004 genome of *Cr. hominis* is not mentioned (a third strain, *Cr. muris*, has since been sequenced). Advanced draft genomes appearing in 2005–2007 but not covered in this book include *Chlamydomonas reinhardtii*, *Giardia lamblia*, two strains of *Ostreococcus tauri*, *Paramecium tetraurelia*, *Phaeodactylum tricornutum*, *Tetrahymena thermophila*, *Theileria annulata* and *Th. parva*, four strains of *Toxoplasma gondii*, *Trichomonas vaginalis*, and the nucleomorph of *Hemiselmis rufescens*. As I write, sequencing and/or annotation continues on *Antonospora locustae*, *Aureococcus anophagefferens*, *Ectocarpus siliculosus*, *Enterozoocoon beineusi*, *Galdieria sulphuraria*, *Oxytricha trifallax* (now *Sterkiella histriomuscorum*) and *Volvox carterae*, while proposals are afoot for other large-scale projects including the genome of *Chondrus crispus*.

In a brief but synthetic introduction, the editors state their intention to “convey a basic understanding of geno-

mics and evolution in microbial eukaryotes to a wide scientific audience". They envision genomics as revolutionising our understanding of not only the physiology, but also the systematics and ecology, of eukaryotes, and see its application to the morphologically simpler representatives as potentially challenging the principles of genome evolution that until now have been based on a few plant and animal lineages. Fifteen chapters are arranged in three sections: an overview of current perspectives on high-level groupings of protists (one chapter); evolutionary genomics of selected taxa or groupings (seven chapters); and analyses of complete genomes (seven chapters, one of which reviews genomic strategies and technologies).

The overview chapter collects many of the better-known eukaryotic lineages into six high-level groupings for which some consensus exists on the basis of inferred molecular-phylogenetic trees, shared morphological (often ultrastructural) features, and/or discrete molecular characters. These groupings are intentionally informal, and the authors are careful to recognise both the provisional state of some of the underlying interpretation, and the great diversity of forms that as yet do not fit into this framework. Judging from the coverage of this book, *microbial eukaryotes* largely excludes two of these six groupings: *plantae/archaeplastida* (*Chlamydomonas* and *Cyanidioschyzon* are unmentioned) and most *opisthokonts* (fungal genomes are excluded although, anomalously, microsporidia are accorded a chapter).

The remaining four high-level groupings, *ramicristates/amoebozoa*, *stramenopiles* and *alveolates*, *rhizaria*, and *excavates*, do not map neatly onto the subsequent fourteen chapters, some of which cut across groups and/or address evolutionary themes including photosynthetic organelles, anaerobes, gene transfer, and remnant nuclear genomes. Among these four groupings *rhizaria* is the least-characterised genomically, and is represented here only by a fascinating if largely pre-genomic chapter on diversity among foraminifera, and by the *Bigelowiella* nucleomorph.

Part II treats evolutionary genomics. As presented by these seven chapters, *evolutionary genomics* encompasses features of genome architecture, chromosome structure, gene content variation, gene family duplication and specialisation, serial endosymbiosis, and gene transfer both between lineages and within the eukaryotic cell. Larger themes include metabolic adaptation, establishment and maintenance of host-parasite systems, genome fusion as an evolutionary mechanism, and the mechanics of genome duality in ciliates. This well-considered selection would have been further enriched by greater emphasis on within-species diversity, e.g., the spectacular chromosomal diversity in *Giardia*, or SNP and haplotype maps of *Plasmodium* or trypanosomes.

The seven chapters in Part III present analyses of complete genomes: trypanosomes, *Entamoeba*, microspori-

dia, nucleomorphs, diatoms and *Dictyostelium* (the *Plasmodium* genomes appear in Part II). In mid-2005 the three trypanosome genomes were the best-assembled and annotated, and the corresponding chapter is packed with information on genome structure, polycistronic gene clusters, telomeres, retro-elements, haplotype information, metabolism, cell-cycle signalling molecules and cross-species comparisons. Gene inventories of *Entamoeba* reveal its metabolic, transport and environmental sensing capacity, repetitive elements, laterally transferred genes and mitosome-targeted gene products. The *Thalassiosira* genome shows two haplotypes, a moderately high level of sequence polymorphism, and relatively few repetitive elements; gene inventories related to silica, nitrogen and carbon metabolism are discussed, while comparison with the *Phaeodactylum* EST library allows preliminary comments on gene diversity between the two diatoms. Among gene inventories for the *Dictyostelium* genome we discover a wealth of G-protein-coupled receptors, orphan receptors, receptor and other kinases, transcriptional activators (although a low frequency of transcription factors), transmembrane-domain proteins, and diverse adhesion-related proteins – many of the parts needed for multicellularity. Much more of course remains to be discovered, not least about the structure and regulation of transcriptional clusters and developmental switches.

The genomic target continues to move. More and more individual genomes and transcript libraries are being sequenced, although increasingly with next-generation technologies not anticipated by this book. Community sequencing (metagenomics) has so far focused on prokaryotes and viruses, but will sooner rather than later be applied to populations of phytoplankton or heterotrophic flagellates. Genomic diversity within populations or species, and genome-wide analyses of transcriptional potential (epigenomics), feature prominently in policy roadmaps of major funding agencies and will soon impact on microbial genomics either directly, or via broad-scale screens for determinants of host resistance to pathogens. In the computational realm, it is increasingly possible to infer novel metabolic pathways and regulatory networks from a combination of sequence and expression data. Too few of these chapters hint at the exciting science and applications that genome-scale technologies have in store for us in the very near future.

This highly readable book is accessible to the molecularly conversant biologist, and represents a very useful first step toward an integrated dynamics genomics of microbial eukaryotes. Many more steps lie ahead.

#### Mark Ragan

Genomics and Computational Biology, Institute for Molecular Bioscience, The University of Queensland, Brisbane, St Lucia QLD 4072, Australia, e-mail: m.ragan@imb.uq.edu.au